



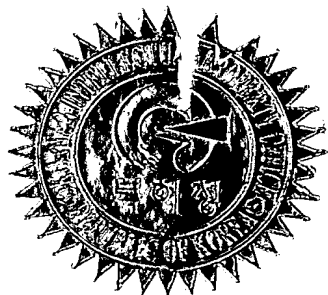
별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

출원 번호 : 10-2003-0095258  
Application Number

출원 년 월 일 : 2003년 12월 23일  
Date of Application DEC 23, 2003

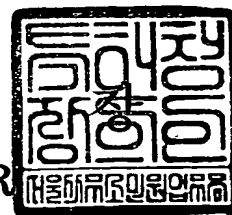
출원인 : 한국전자통신연구원  
Applicant(s) Electronics and Telecommunications Research Insti



2003 년 12 월 30 일

특 허 청

COMMISSIONER





## 【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0013
【제출일자】	2003.12.23
【발명의 명칭】	보이스엑스엠엘 문서를 엑스에이치티엠엘플러스보이스 문서로 변환하기 위한 방법 및 이를 이용한 멀티모달 서비스 시스템
【발명의 영문명칭】	Method for VoiceXML to XHTML+Voice Conversion and Multimodal Service System using the same
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【성명】	권태복
【대리인코드】	9-2001-000347-1
【포괄위임등록번호】	2001-057650-1
【대리인】	
【성명】	이화익
【대리인코드】	9-1998-000417-9
【포괄위임등록번호】	1999-021997-1
【발명자】	
【성명의 국문표기】	김지은
【성명의 영문표기】	KIM, Ji-Eun
【주민등록번호】	761125-2623811
【우편번호】	305-503
【주소】	대전광역시 유성구 송강동 한솔아파트 202동 402호
【국적】	KR
【발명자】	
【성명의 국문표기】	박지은
【성명의 영문표기】	PARK, Ji-Eun
【주민등록번호】	690714-2721618

【우편번호】 305-721  
 【주소】 대전광역시 유성구 신성동 럭키하나아파트 109동 302호  
 【국적】 KR

## 【발명자】

【성명의 국문표기】 박준석  
 【성명의 영문표기】 PARK, Jun-Seok  
 【주민등록번호】 621218-1148112

【우편번호】 305-755  
 【주소】 대전광역시 유성구 어은동 한빛아파트 112동 504호  
 【국적】 KR

## 【발명자】

【성명의 국문표기】 한동원  
 【성명의 영문표기】 HAN, Dong-Won  
 【주민등록번호】 580916-1074415  
 【우편번호】 302-791  
 【주소】 대전광역시 서구 월평동 누리아파트 107동 1008호  
 【국적】 KR

## 【공지예외적용대상증명서류의 내용】

【공개형태】 학술단체 서면발표  
 【공개일자】 2003.12.09

## 【공지예외적용대상증명서류의 내용】

【공개형태】 학술단체 서면발표  
 【공개일자】 2003.12.21

## 【심사청구】

청구

## 【취지】

특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인  
 권태복 (인) 대리인  
 이화익 (인)

## 【수수료】

【기본출원료】	20	면	29,000	원
【가산출원료】	6	면	6,000	원
【우선권주장료】	0	건	0	원
【심사청구료】	9	항	397,000	원



1020030095258

출력 일자: 2004/1/2

【합계】	432,000 원
【감면사유】	정부출연연구기관
【감면후 수수료】	216,000 원
【기술이전】	
【기술양도】	희망
【실시권 허여】	희망
【기술지도】	희망
【첨부서류】	1. 요약서·명세서(도면)_1통 2. 공지에외적용대상(신규성장실의예 외, 출원시의특례)규정을 적용받 기 위한 증명서류_2통

{

**【요약서】****【요약】**

본 발명은 VoiceXML기반 음성 서비스를 XHTML(eXtensible HyperText Markup Language) 기반 웹 인터페이스와 VoiceXML 기반 음성 인터페이스를 지원하는 XHTML+ Voice기반 멀티모달 서비스로 변환하는 방법 및 시스템에 관한 것이다. 본 발명의 변환 방법은 VoiceXML 문서를 파싱한 트리를 상위 태그부터 하위 태그까지 모두 검색하면서 XHTML+Voice 트리를 초기화하는 단계; 태그를 확인하여 태그가 <menu>이면 XHTML의 <a> 태그로 변환하는 단계; 태그를 확인하여 태그가 <grammar>이면 XHTML의 <input type=radio> 태그로 변환하는 단계; 및 태그를 확인하여 태그가 <form>이면 XHTML의 <form>을 XHTML+Voice 트리에 추가한 후 <form> 태그를 처리하는 단계를 포함한다. 또한, 본 발명에 따른 시스템은 프록시 서버(Proxy Server)와 같은 별도의 외부 시스템이나 통상의 사용자 디바이스의 XHTML+Voice 브라우저에 상기 변환 방법을 구현한 트랜스코더(Transcoder) 또는 트랜스코더의 부분 모듈(Module)을 탑재하는 모습으로 나타날 수 있다.

**【대표도】**

도 4

**【색인어】**

VoiceXML, multimodal, XHTML+Voice, XHTML+Voice, 마크업 언어, 변환

**【명세서】****【발명의 명칭】**

보이스엑스엠엘 문서를 엑스에이치티엠엘플러스보이스 문서로 변환하기 위한 방법 및 이를 이용한 멀티모달 서비스 시스템{Method for VoiceXML to XHTML+Voice Conversion and Multimodal Service System using the same}

**【도면의 간단한 설명】**

도 1은 전화망상의 VoiceXML기반 음성 서비스 방식을 설명하기 위한 도면.

도 2는 본 발명에 따른 트랜스코더가 프록시 서버에 구현된 경우를 도시한 구성 블록도.

도 3은 본 발명에 따른 트랜스코더의 한 모듈인 VoiceXML-to-XHTML+Voice 변환기가 XHTML+Voice 브라우저에 내장된 경우를 도시한 구성 블록도.

도 4는 본 발명에 따른 트랜스코더의 한 모듈인 VoiceXML-to-XHTML+Voice 변환기의 알고리즘을 도시한 순서도.

도 5는 본 발명에 따른 변환 전의 예제 음성 시나리오와 변환 후 XHTML+Voice 브라우저에서 실행한 화면을 도시한 도면.

도 6은 도 5의 예제 음성 시나리오의 VoiceXML 문서 구조를 도시한 도면.

도 7은 도 5의 예제 음성 시나리오의 VoiceXML 트리와 본 발명에 따라 변환하여 생성된 XHTML+Voice 트리를 도시한 도면.

도 8은 도 7의 XHTML+Voice 트리에서 생성된 XHTML+Voice 문서 구조를 도시한 도면.

<도면의 주요부분에 대한 부호의 설명>

210: 사용자	211: XHTML+Voice 브라우저
220: 프록시 서버	230: 트랜스코더
231: VoiceXML 파서	232: VoiceXML-to-XHTML+Voice 변환기
233: XHTML+Voice 문서 생성기	240: 웹 서버
242: VoiceXML 응용	

**【발명의 상세한 설명】**

**【발명의 목적】**

**【발명이 속하는 기술분야 및 그 분야의 종래기술】**

<15>        본 발명은 VoiceXML(Voice eXtensible Markup Language)기반 음성 서비스를 XHTML(eXtensible HyperText Markup Language)기반 웹 인터페이스와 VoiceXML 기반 음성 인터페이스를 지원하는 XHTML+Voice 기반 멀티모달 서비스(Multimodal Service)로 변환하는 방법 및 시스템에 관한 것이다.

<16>        일반적으로, VoiceXML은 음성 인식(Speech Recognition) 및 음성 합성(Text to Speech) 기술과 컴퓨터 전화통합(CTI:Computer Telephony Integration) 기술에 웹 정보 처리 기술을 접목한 음성 대화 시나리오 저작 표준 언어이다. 즉, VoiceXML은 XML에 기반을 둔 마크업 언어(Markup language)로서 유선, 이동전화를 이용해 인터넷 정보를 음성으로 검색하고 들을 수 있는 음성 다이얼로그(spoken

dialog)를 정의하는데 사용되는 언어이다. VoiceXML 문서는 노트북(Notebook), 피시/Desktop PC) 등 인터넷 접속장치 없이 유선, 이동전화를 통해서 인터넷상의 전자우편, 날씨정보, 교통 정보 등의 내용을 검색할 수 있으며 웹 페이지(Web page)의 내용을 음성으로 제공할 수 있다.

<17> 따라서, VoiceXML은 웹을 통해 실시간으로 서비스 창출 및 유지 보수를 할 수 있어 기존 자동응답서비스(ARS:Automatic Response Services)와, 상호응답(IVR:Interactive Voice Response) 등의 대화형 음성 서비스 시스템을 대체할 차세대 음성 서비스의 핵심 기술로 인정 받고 있다.

<18> 도 1은 전화망상의 VoiceXML 기반 음성 서비스 방식을 설명하기 위한 도면으로서, 사용자(102-1,102-2), 공중전화망(PSTN: 104), IVR(106), 인터넷(108), 음성 게이트웨이(110), 웹 서버(120)가 도시되어 있다. 사용자(102-1)는 전화기나 이동전화기를 이용하여 음성 웹 서비스를 사용할 수 있고, 사용자(102-2)는 PC를 통해 웹 서버(120)에 접속하여 통상의 웹서비스를 사용할 수 있다. 웹 서버(120)는 통상의 웹 페이지와 함께 VoiceXML 응용(122)을 구비하여 인터넷(108)을 통해 사용자(102-2)에게 웹 페이지를 제공함과 아울러 음성 게이트웨이(110)의 HTTP 요구에 따라 VoiceXML 문서를 제공한다. 음성 게이트웨이(110)는 VoiceXML 브라우저(112)와 음성인식/합성기(114), 스크립트 엔진(116) 등을 구비하여 사용자(102-1)의 요구에 따라 웹 서버(120)에 HTTP 요구를 통해 음성 웹 문서를 요구하고, VoiceXML문서를 받으면 VoiceXML 브라우저(112)를 통해 이를 실행한 후 음성인식/합성기(114)로 공중전화망(104)을 통해 사용자에게 전달한다.

<19> 이와 같은 전화망을 통한 음성 웹 서비스 동작은 다음과 같다.

<20> 먼저, 사용자(102-1)는 유선 전화기 또는 이동 통신 단말기를 통하여 대표번호로 음성 게이트웨이(Voice Gateway)(110)에 접속한다. 음성 게이트웨이(110)의 VoiceXML 브라우저



(112)는 VoiceXML 문서를 웹 서버(120)에 요청한다. 웹 서버(120)는 해당 VoiceXML 문서를 전송한다. 그리고 음성 게이트웨이의 VoiceXML 브라우저(112)는 전송된 VoiceXML 문서를 해석하여 실행한 음성 출력을 전화망(104)을 통하여 사용자(102-1)에게 제공한다.

<21> 한편, 현재 많은 분야(예를 들어, 증권, 카드, 유통 등)에서 다양한 서비스를 제공하는 VoiceXML 기반 음성 서비스를 PDA나 스마트폰, PC 등에서 인터넷과 브라우저를 통해 서비스 받고자 한다면, VoiceXML에 대한 소정의 변환 과정이 필요하다. 이때 인터넷과 브라우저를 통해 서비스 받는다는 것은 디바이스의 성격이 음성 이외의 인터페이스도 제동됨을 의미하기 때문에 변환 과정에서 사용자 인터페이스의 변화도 고려되어야 한다.

<22> 이러한 요구를 충족시킬 수 있는 마크업 언어로 XHTML+Voice가 있다. 이는 XHTML기반 웹 서비스와 VoiceXML(VoiceXML 2.0의 서브셋)기반 음성 서비스를 결합한 멀티모달 웹 서비스를 개발하기 위한 목적으로 제안되었다. XHTML+Voice 문서 작성은 기존의 XHTML 문서 작성 방법 및 VoiceXML 문서 작성 방법과 비슷하지만 음성과 연관되는 태그(tag)들은 XML 이벤트와 XHTML+Voice 이벤트를 통해 연동되도록 하고 있다. 따라서 현재 제공되는 VoiceXML 기반 음성 서비스를 PDA나 스마트폰, PC 등에서 인터넷과 브라우저를 통해 멀티모달 서비스로 받고자 한다면, 기존의 VoiceXML 문서를 XHTML+Voice 문서로 변환하는 과정이 필요하다.

<23> 그런데 XHTML+Voice에서 지원되는 음성(Voice) 태그는 VoiceXML의 서브셋이기 때문에 XHTML+Voice 태그와 VoiceXML 태그는 1:1 매칭으로 변환할 수 없는 문제점이 있다. 또한 VoiceXML 문서는 순차적 구조로서 순차적 입력만을 처리하는데, 이를 병렬 구조를 갖는 XHTML+Voice 문서로 적절히 변경하여야 한다.

【발명이 이루고자 하는 기술적 과제】

- <24> 본 발명은 상기와 같은 문제점을 해결하기 위해, 소정의 변환 알고리즘으로 VoiceXML 문서를 XHTML+Voice 문서로 변환하는 변환방법 및 이를 이용한 멀티모달 시스템 구성에 그 목적이 있다.
- <25> 상기와 같은 목적을 달성하기 위하여 본 발명의 변환방법은 VoiceXML 문서를 파싱하여 생성된 VoiceXML 트리를 XHTML+Voice 트리로 변환하는 방법에 있어서, 상기 VoiceXML 트리를 상위 태그부터 하위 태그까지 모두 검색하면서, XHTML+Voice 트리를 초기화하는 단계; 태그를 확인하여 태그가 <menu>이면 XHTML+Voice의 <a> 태그로 변환하는 단계; 태그를 확인하여 태그가 <grammar>이면 XHTML+Voice의 <input type=radio> 태그로 변환하는 단계; 및, 태그를 확인하여 태그가 <form>이면 XHTML+Voice의 <form>을 XHTML+Voice 트리에 추가한 후 <form> 태그를 처리하는 단계를 포함하는 것을 특징으로 한다.
- <26> 또한, 상기와 같은 목적을 달성하기 위하여 본 발명의 멀티모달 서비스 방법은 통상의 XHTML+Voice 브라우저가 탑재된 사용자 단말기와, 프록시 서버와, VoiceXML 문서를 제공하는 웹 서버를 포함하는 시스템에 있어서, 사용자 단말기에서 XHTML+Voice 브라우저를 구동하여 HTTP 요구(request)를 통해 웹 서버로 VoiceXML 문서를 요청하는 단계; 상기 웹 서버가 VoiceXML 문서를 상기 프록시 서버에 전송하는 단계; 상기 프록시 서버에 탑재된 VoiceXML 파서는 전송받은 VoiceXML문서를 트리 구조로 구성하여 VoiceXML-to-XHTML+Voice 변환기로 전달하는 단계; VoiceXML-to-XHTML+Voice 변환기는 전송된 VoiceXML 트리를 소정의 알고리즘으로 새로운 XHTML+Voice 트리로 변환하여 XHTML+Voice 생성기로 전달하는 단계; XHTML+Voice 문서 생성기는 XHTML+Voice 트리를 입력받아 XHTML+Voice 문서를 생성하여 상기 XHTML+Voice 브라우저

저로 전송하는 단계; 및, 상기 사용자의 XHTML+Voice 브라우저가 XHTML+Voice 문서를 해석하여 실행한 후 음성 및 그래픽을 출력하는 단계를 포함하는 것을 특징으로 한다.

### 【발명의 구성 및 작용】

<27> 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예를 자세히 설명하기로 한다.

<28> 본 발명에 따라 VoiceXML 문서를 XHTML+Voice 문서로 변환하는 모듈(이하, 'VoiceXML-to-XHTML+Voice 변환기'라 한다)은 사용자 디바이스의 XHTML+Voice 브라우저에 내장될 수 있다(제2 실시예). 만약, 본 발명의 VoiceXML-to-XHTML+Voice 변환기를 탑재한 XHTML+Voice 브라우저를 사용하지 않는 사용자 디바이스에서 음성 서비스를 제공받고자 한다면, 본 발명의 VoiceXML-to-XHTML+Voice 변환기를 구비한 트랜스코더가 동작하는 프록시 서버(Proxy Server)를 통해 변환된 XHTML+Voice 문서를 전송받아야 한다(제1 실시예).

<29> [제1 실시예]

<30> 도 2는 본 발명의 트랜스코더가 프록시 서버에 내장된 경우로서, 사용자(210)와 프록시 서버(220), 및 웹 서버(240)의 관계가 도시되어 있다. 사용자(210)는 XHTML+Voice 브라우저(211)와 음성인식기(215), 음성합성기(216), 스크립트 엔진(Script Engine; 217)을 포함하고 있고, 프록시 서버(Proxy Server; 220)는 트랜스코더(Transcoder; 230)를 구비하고 있는데 트랜스코더(230)는 VoiceXML 파서(231)와 VoiceXML-to-XHTML+Voice 변환기(232), XHTML+Voice 문서 생성기(233)로 이루어진다. 그리고 웹 서버(Web Server; 240)는 VoiceXML 응용(Application; 242)을 구비하고 있다.

- <31> 도 2를 참조하면, 통상의 XHTML+Voice 브라우저(211)는 XHTML문서를 XHTML 트리로 구성하는 XHTML 파서(Parser; 213), VoiceXML 문서를 VoiceXML 트리로 구성하는 VoiceXML 파서(212), 각 트리를 실행하여 인터랙션하는 XHTML+Voice 렌더러(Render; 214)로 구성된다. 이러한 XHTML+Voice 브라우저(211)는 스크립트 엔진(217)을 이용하여 ECMA 스크립트를 처리하고, 음성 합성기(216)를 이용하여 음성을 출력하며, 음성 인식기(215)를 이용하여 음성 입력을 처리한다. 또한 XHTML+Voice 브라우저(211)는 텍스트 입력(터치 스크린, 하드웨어 키보드를 포함)도 처리한다.
- <32> 서비스 사업자는 음성 서비스를 저작하여 웹 서버(240) 등을 통해 서비스하는데, 웹 서버(240)는 VoiceXML 응용(242)을 통해 프록시 서버(220)로부터 HTTP 요구(request)가 수신되면 해당 VoiceXML 문서를 전송해 준다.
- <33> 프록시 서버(220)는 본 발명에 따라 VoiceXML문서를 XHTML+Voice 문서로 변환해주는 트랜스코더(230)를 포함하는데, 본 발명의 트랜스코더(230)는 VoiceXML 트리를 생성하는 VoiceXML 파서(231)와 소정의 변환 알고리즘을 구현한 VoiceXML-to-XHTML+Voice 변환기(232)와 XHTML+Voice 트리를 XHTML+Voice 문서로 변환하기 위한 XHTML+Voice 문서 생성기(Generator)(233)로 구성된다.
- <34> 이와 같은 본 발명의 트랜스코더(230)를 이용하여 통상의 XHTML+Voice 브라우저(211)를 사용하는 사용자(210)에게 멀티모달 서비스를 제공하는 절차는 다음과 같다.
- <35> 사용자(210)는 PDA 또는 스마트 폰과 같은 단말기를 통하여 XHTML+Voice 브라우저(211)를 구동한다. 이어 HTTP 요구(request)을 통해 웹 서버(240)로 VoiceXML 문서를 요청한다. 웹 서버(240)는 VoiceXML 문서를 프록시 서버(220)로 전송한다.

- <36>       프록시 서버(220)에 탑재된 VoiceXML 파서(231)는 전송받은 VoiceXML문서를 트리 구조로 구성하고, 생성된 VoiceXML 트리를 VoiceXML-to-X+V 변환기(232)로 전달한다.
- <37>       VoiceXML-to-XHTML+Voice 변환기(232)는 전송된 VoiceXML 트리를 소정의 알고리즘으로 새로운 XHTML+Voice 트리로 변환하고, 변환된 XHTML+Voice 트리를 XHTML+Voice 문서 생성기(233)로 전달한다. XHTML+Voice 문서 생성기(233)는 XHTML+Voice 트리를 입력받아 XHTML+Voice 문서를 생성하고, XHTML+Voice 브라우저(211)로 전송한다.
- <38>       이에 따라 사용자(210)의 XHTML+Voice 브라우저(211)는 XHTML+Voice 문서를 해석하여 실행한 후 음성 및 그래픽으로 출력한다.
- <39>       [제2 실시예]
- <40>       도 3은 본 발명의 VoiceXML-to-XHTML+Voice 변환기가 XHTML+Voice 브라우저에 내장된 경우를 도시한 도면으로서, 사용자(310)와 웹 서버(240)의 관계가 도시되어 있다.
- <41>       도 3을 참조하면, 사용자(310)의 단말기에는 XHTML+Voice 브라우저(320)와 음성인식/합성기(TTS & SRS; 332), 스크립트 엔진(Script Engine; 334)이 탑재되어 있고, XHTML+Voice 브라우저(320)는 VoiceXML 파서(321), VoiceXML-to-XHTML+Voice 변환기(322), XHTML+Voice 렌더러(323)를 포함하고 있다. VoiceXML 파서(321)는 VoiceXML문서에서 VoiceXML 트리를 생성하고, VoiceXML-to-XHTML+Voice 변환기(322)는 소정의 변환 알고리즘에 따라 VoiceXML 트리에서 XHTML+Voice 트리를 생성하고, XHTML+Voice 렌더러(Render; 323) XHTML+Voice 트리를 실행하여 음성인식/합성기(332)를 통해 음성으로 출력한다. 그리고 스크립트 엔진(334)은 ECMA 스크립트를 처리한다.



- <42> 이러한 본 발명의 XHTML+Voice 브라우저(320)를 이용하여 멀티모달 서비스를 제공하는 절차는 다음과 같다.
- <43> 사용자(310)는 PDA 또는 스마트 폰과 같은 단말기를 통하여 XHTML+Voice 브라우저(320)를 구동한다. XHTML+Voice 브라우저(320)는 HTTP 호출을 통해 웹 서버(240)로 VoiceXML 문서를 요청한다. 이에 따라 웹 서버의 VoiceXML 응용(242)은 해당 VoiceXML 문서를 XHTML+Voice 브라우저(320)로 전송한다.
- <44> XHTML+Voice 브라우저(320)의 VoiceXML 파서(321)는 전송 받은 VoiceXML 문서를 트리 구조로 구성하고, 생성된 VoiceXML 트리를 VoiceXML-to-XHTML+Voice 변환기(322)로 전달한다. VoiceXML-to-XHTML+Voice 변환기(322)는 전송된 VoiceXML 트리를 소정의 알고리즘으로 새로운 XHTML+Voice 트리로 변환하고 이를 XHTML+Voice 랜더러(323)로 전달한다. XHTML+Voice 랜더러(323)는 XHTML+Voice 트리를 해석하여 실행한 후 음성 및 그래픽으로 출력한다.
- <45> 도 4는 본 발명에 따른 VoiceXML-to-XHTML+Voice 변환기의 변환 알고리즘을 도시한 순서도이다.
- <46> 도 4를 참조하면, VoiceXML 트리를 상위 태그부터 하위 태그까지 모두 검색하면서, XHTML+Voice 트리를 초기화 한다(401,402). 이중 메인 다이얼로그는 새로 생성되는 XHTML 트리이다.
- <47> 태그를 확인하여 처음 태그가 <form>, <menu>, <grammar>인지 확인한다(403).
- <48> 만일, 태그가 <menu>이면 XHTML의 <a> 태그로 변환한 후 VoiceXML 트리를 삭제한다(404~406).

- <49> 만일, 태그가 <grammar>이면 XHTML의 <input type=radio> 태그로 변환한 후 이벤트/핸들러를 정의한다(407~409).
- <50> 만일, 태그가 <form>이면 XHTML의 <form>을 XHTML 트리에 추가한다(411). 하나의 <form> 태그에 속하는 <block>과 <prompt> 태그는 PC DATA이면 XHTML의 <p> 태그로 변환한 후 이벤트/핸들러를 정의한다(418~421).
- <51> 하나의 <form>과 <field> 태그에 속한 <prompt> 태그는 XHTML의 <label> 태그로 변환하고, 하위 태그로 <input type=text> 태그를 생성한 후 이벤트/핸들러를 정의하고 VoiceXML을 수정한다(412~417).
- <52> 하나의 <form>과 <field> 또는 <block>에 속하는 <submit> 태그는 XHTML의 <input type=submit> 태그로 변환한 후 이벤트/핸들러를 정의하고, VoiceXML을 수정한다(422~425). 이와 같이 모든 과정마다 적절한 이벤트가 추가되어야 하고, 때로는 대상 트리인 VoiceXML 트리를 수정하거나 삭제하여야 한다.
- <53> 이와 같은 본 발명의 변환 알고리즘을 이해하기 쉽도록 하나의 예제를 통해 확인해보면 다음과 같다.
- <54> 도 5는 본 발명에 따른 변환 전의 예제 음성 시나리오와, 변환 후에 XHTML+Voice 브라우저에서 실행한 화면의 예를 도시한 도면이다.
- <55> 도 5를 참조하면, 변환 전의 예제 음성 시나리오(510)는 항공 예약과 관련된 시나리오로서, 사용자는 PDA나 스마트폰으로 인터넷을 통해 제공되는 음성 서비스 중 하나인 비행기 예약 서비스를 받고자 한다. 서비스 사업자가 제공하는 비행기 예약 서비스의 시나리오(510)는 사용자 이름("What is your name"), 출발지역("The city of your departure"), 도착지역("The



city of your destination"), 출발 예정일("The date of your departure") 등을 차례로 입력 받아 처리하는 구성으로 되어 있다.

<56> 그리고 이와 같은 시나리오를 갖는 VoiceXML 문서는 본 발명에 따라 변환된 후 XHTML+Voice 브라우저에서 실행되어 도면의 오른쪽과 같이 화면(520)으로 나타난다.

<57> XHTML+Voice 브라우저 화면(520)은 기본적으로 음성 사용 모드를 지원하기 때문에 사용자가 임의의 입력 창을 선택(click & focus)하면, 해당 질문을 음성으로 읽어주고 적절한 값을 음성으로 입력받기 위해 대기한다. 만약, 음성 취소(voice\_cancel) 버튼(522)을 클릭하여 음성 취소 모드를 선택한다면, 사용자는 텍스트만을 이용하여 입력하여야 한다. 모든 입력이 끝나면, 처리(submit) 버튼(521)을 클릭하여 입력 내용을 다음 단계의 응용 프로그램에게 전달한다

<58> 도 6은 도 5의 예제 음성 시나리오의 VoiceXML 문서 구조를 보여주는 도면으로서, 예제 음성 시나리오의 VoiceXML 문서는 메인 다이얼로그인 app.vxml 문서(610)와, 서브 다이얼로그인 sub\_app.vxml 문서(620)로 이루어진다.

<59> 도 6을 참조하면, 메인 다이얼로그(Main-dialog)인 app.vxml 문서(610)는 하나의 <form>에 <field a>(611), <subdialog>(612), <field b>(613), <submit>(614) 태그로 구성되어 있다. 서브 다이얼로그(Sub-dialog)인 sub\_app.vxml 문서(620)는 하나의 <form>에 <field c>(621), <field d>(622), <return>(623) 태그로 구성되어 있다. 본 발명의 실시예에서 "Welcome to the Flight Reservation Service" 는 <block> 태그에 속하지만 설명에서는 제외한다.





- <60> 도 7은 도 5의 예제 음성 시나리오의 VoiceXML 트리와 이를 본 발명의 변환 알고리즘을 적용하여 생성한 XHTML+Voice 트리를 도시한 도면이다.
- <61> 도 7을 참조하면, 예제 음성 시나리오의 VoiceXML 트리는 app 트리(710)와 sub\_app 트리(720)로 이루어지고, 본 발명의 변환 알고리즘에 의해 new 트리(730), 변형된 app 트리(710'), 변형된 sub\_app 트리(720')로 생성 또는 변형된 것을 알 수 있다.
- <62> app 트리(710)는 하나의 form이 제1 필드(field), 서브 다이얼로그(subdialog), 제2 필드(field), 블록(block)으로 이루어지고, sub\_app 트리(720)는 하나의 form이 다시 2개의 필드(field)로 이루어진다.
- <63> 도 8은 도 7의 XHTML+Voice 트리에서 생성된 XHTML+Voice 문서 구조를 보여주는 도면이다.
- <64> 도 8을 참조하면, 메인 다이얼로그인 new.vxml 문서(810)는 최상위 태그 <html>에 <head> 태그(820), <body> 태그(830)를 기본 구조로 갖는다.
- <65> <head> 태그(820)는 음성 문서의 <field> 태그와 <body>의 <input> 태그를 동기화(synchronization)(802)하기 위한 <xv:sync> 태그(821)와, 음성 취소 모드를 처리하는 <xv:cancel> 태그(821)로 구성된다.
- <66> <body> 태그(830)는 하나의 <form> 태그에 음성 문서의 <field> 태그에서 변환된 <input type=text a>(831), <input type=text c>(832), <input type=text d>(833), <input type=text b> 태그(834)와 음성 문서의 <submit> 태그에서 변환된 <input type=submit> 태그(835), 그리고 음성 취소 모드를 위한 <input type=reset> 태그(836)로 구성된다.



<67> app.vxml 문서(840)는 <form a>(841)에 <field a>, <form b>(842)에 <field b> 태그를 갖는 서브 다이얼로그로 수정되었다. sub\_app.vxml 문서(850)는 <form c>(851)에 <field c>, <form d>(852)에 <field d>를 갖는 서브 다이얼로그로 수정되었다.

### 【발명의 효과】

<68> 이 상에서 설명한 바와 같이, 본 발명의 VoiceXML-to-XHTML+Voice 변환기 및 이를 포함하는 트랜스코더는, XHTML+Voice 태그와 VoiceXML 태그를 가능한 1:1 매칭으로 변환하지만, 그렇지 못한 호 제어(Call Control) 태그의 경우는 스크립트를 이용하여 시스템을 제어하는 방법이나 응용 프로그램을 이용하는 방법 혹은 아예 삭제하는 방법을 통해 해결한다. 또한 본 발명의 VoiceXML-to-XHTML+Voice 변환기는 사용자 디바이스에 내장하거나 아니면 트랜스코더를 갖는 프록시 서버와 같은 시스템으로 따로 구축할 수 있도록 하여 사용자 환경에 맞추어 서비스를 제공할 수 있다.

<69> 또한 본 발명에 의하면 서비스 제공자는, 실시간으로 전화망을 위한 VoiceXML 기반 음성 서비스를 인터넷을 위한 XHTML+Voice 멀티모달 서비스로 자동 변환함으로써, 기존의 VoiceXML 기반 음성 서비스를 활용하여 XHTML+Voice 기반의 멀티모달 서비스를 쉽게 구축할 수 있다. 즉, PDA나 스마트폰과 같은 지능 정보형 디바이스를 위한 서비스를 다시 개발하지 않아도 저렴한 비용으로 멀티모달 서비스를 구축할 수 있다. 그리고 VoiceXML 기반 음성 서비스의 유지 보수만으로 멀티모달 서비스를 위한 유지 보수는 자동 수행되는 효과가 발행하여, 멀티모달 서비스를 위한 별도의 유지보수 비용이 거의 발생하지 않는다.

- <70> 그리고 본 발명에 의하면 서비스 이용자는, 인터넷으로 음성 서비스를 이용하는데 있어서, 싱글모달이 아닌 멀티모달 인터페이스로 인터랙션(Interaction)하며, 순차적이 아닌 병렬적으로 서비스를 제어하고, 모드 스위칭(음성 사용여부를 선택)을 통해 사용하고자 하는 모드를 선택할 수 있으며, 결과적으로 불필요한 인터랙션(User Overexertion)을 줄여 보다 정확하고 효율적으로 음성 서비스를 이용할 수 있다.
- <71> 한편, 본 발명에 적합한 음성 서비스로는, 날씨/뉴스/증권/교통 정보 등의 실시간 정보 안내 서비스, 음식의 요리법이나 응급환자의 응급처치와 같은 절차적 내용을 갖는 서비스, 여론조사/시청률 조사/소비자 정보 조사 등 각종 갤럽 조사 서비스, 잔액조회/각종 은행 상품 정보 검색이 가능한 은행 서비스 등이 있다.
- <72> 상기에서는 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.



【특허청구범위】

【청구항 1】

VoiceXML 문서를 파싱하여 생성된 VoiceXML 트리를 XHTML+Voice 트리로 변환하는 방법에 있어서,

(a) 상기 VoiceXML 트리를 상위 태그부터 하위 태그까지 모두 검색하면서, XHTML+Voice 트리를 초기화하는 단계;

(b) 태그를 확인하여 태그가 <menu>이면 XHTML의 <a> 태그로 변환하는 단계;

(c) 태그를 확인하여 태그가 <grammar>이면 XHTML의 <input type=radio> 태그로 변환하는 단계; 및

(d) 태그를 확인하여 태그가 <form>이면 XHTML의 <form>을 XHTML 트리에 추가한 후 <form> 태그를 처리하는 단계를 포함하는 것을 특징으로 하는

VoiceXML 문서를 XHTML+Voice 문서로 변환하는 방법.

【청구항 2】

제1항에 있어서, 상기 (d) 단계는

하나의 <form>에 속하는 <block>과 <prompt> 태그는 XHTML의 <p> 태그로 변환하는 단계;

하나의 <form>과 <field> 태그에 속한 <prompt> 태그는 XHTML의 <label> 태그로 변환하는 단계; 및



하나의 <form>과 <field> 또는 <block>에 속하는 <submit> 태그는 XHTML의 <input type=submit> 태그로 변환하는 단계를 포함하는 것을 특징으로 하는

VoiceXML 문서를 XHTML+Voice 문서로 변환하는 방법.

### 【청구항 3】

제1항 또는 제2항에 있어서, 상기 각 단계는

변환 후 이벤트/핸들러를 정의하거나 VoiceXML을 수정 혹은 삭제하는 것을 특징으로 하는

VoiceXML 문서를 XHTML+Voice 문서로 변환하는 방법.

### 【청구항 4】

통상의 XHTML+Voice 브라우저가 탑재된 사용자 단말기와, 프록시 서버와, VoiceXML 문서를 제공하는 웹 서버를 포함하는 시스템에서 VoiceXML 문서를 XHTML+Voice 문서로 변환하는 것을 이용한 멀티모달 서비스 방법에 있어서,

사용자 단말기에서 XHTML+Voice 브라우저를 구동하여 HTTP 요구(request)를 통해 웹 서버로 VoiceXML 문서를 요청하는 단계;

상기 웹 서버가 VoiceXML 문서를 상기 프록시 서버에 전송하는 단계;

상기 프록시 서버에 탑재된 VoiceXML 파서는 전송받은 VoiceXML문서를 트리 구조로 구성하여 VoiceXML-to-XHTML+Voice 변환기로 전달하는 단계;

VoiceXML-to-XHTML+Voice 변환기는 전송된 VoiceXML 트리를 소정의 알고리즘으로 새로운 XHTML+Voice 트리로 변환하여 XHTML+Voice 문서 생성기로 전달하는 단계;

XHTML+Voice 문서 생성기는 XHTML+Voice 트리를 입력받아 XHTML+Voice 문서를 생성하여 상기 XHTML+Voice 브라우저로 전송하는 단계; 및

상기 사용자의 XHTML+Voice 브라우저가 XHTML+Voice 문서를 해석하여 실행한 후 음성 및 그래픽을 출력하는 단계를 포함하는 것을 특징으로 하는

VoiceXML 문서를 XHTML+Voice 문서로 변환하는 것을 이용한 멀티모달 서비스 방법.

#### 【청구항 5】

VoiceXML-to-XHTML+Voice 변환기를 갖는 XHTML+Voice 브라우저가 탑재된 사용자 단말기와 VoiceXML 문서를 제공하는 웹 서버를 포함하는 시스템에서 VoiceXML 문서를 XHTML+Voice 문서로 변환하는 것을 이용한 멀티모달 서비스 방법에 있어서,

사용자 단말기를 통하여 XHTML+Voice 브라우저를 구동하여 HTTP 호출을 통해 상기 웹 서버로 VoiceXML 문서를 요청하는 단계;

상기 웹 서버가 해당 VoiceXML 문서를 XHTML+Voice 브라우저로 전송하는 단계;

XHTML+Voice 브라우저의 VoiceXML 파서가 전송 받은 VoiceXML 문서를 트리 구조로 구성하여 VoiceXML-to-XHTML+Voice 변환기로 전달하는 단계;

상기 VoiceXML-to-XHTML+Voice 변환기가 전송된 VoiceXML 트리를 소정의 알고리즘으로 새로운 XHTML+Voice 트리로 변환하는 단계; 및



XHTML+Voice 랜더러가 XHTML+Voice 트리를 해석하여 실행한 후 음성 및 그래픽으로 출력하는 단계를 포함하는 것을 특징으로 하는

VoiceXML 문서를 XHTML+Voice 문서로 변환하는 것을 이용한 멀티모달 서비스 방법.

#### 【청구항 6】

XHTML+Voice 브라우저가 탑재된 사용자 단말기와, 프록시 서버와, VoiceXML 문서를 제공하는 웹 서버를 포함하는 시스템에 있어서,

VoiceXML 트리를 생성하는 VoiceXML 파서와,

소정의 변환 알고리즘을 구현한 VoiceXML-to-XHTML+Voice 변환기와,

XHTML+Voice 트리를 XHTML+Voice 문서로 변환하기 위한 XHTML+Voice 문서 생성기(Generator)로 구성된 트랜스코더가 상기 프록시 서버에 탑재된 것을 특징으로 하는 멀티모달 서비스 시스템.

#### 【청구항 7】

XHTML+Voice 브라우저가 탑재된 사용자 단말기와, VoiceXML 문서를 제공하는 웹 서버를 포함하는 시스템에 있어서,

상기 XHTML+Voice 브라우저가

VoiceXML 문서에서 VoiceXML 트리를 생성하는 VoiceXML 파서;

소정의 변환 알고리즘에 따라 VoiceXML 트리에서 XHTML+Voice 트리를 생성하는 VoiceXML-to-XHTML+Voice 변환기; 및



상기 XHTML+Voice 트리를 실행하는 XHTML+Voice 렌더러(Render)를 포함하는 것을 특징으로 하는

멀티모달 서비스 시스템.

#### 【청구항 8】

제7항에 있어서, 상기 XHTML+Voice 브라우저를 통해 서비스 되는 음성 서비스는

멀티모달 서비스로 브라우징되며, 음성 입출력 사용 모드와 음성 입출력 취소 모드 선택이 가능한 것을 특징으로 하는

멀티모달 서비스 시스템.

#### 【청구항 9】

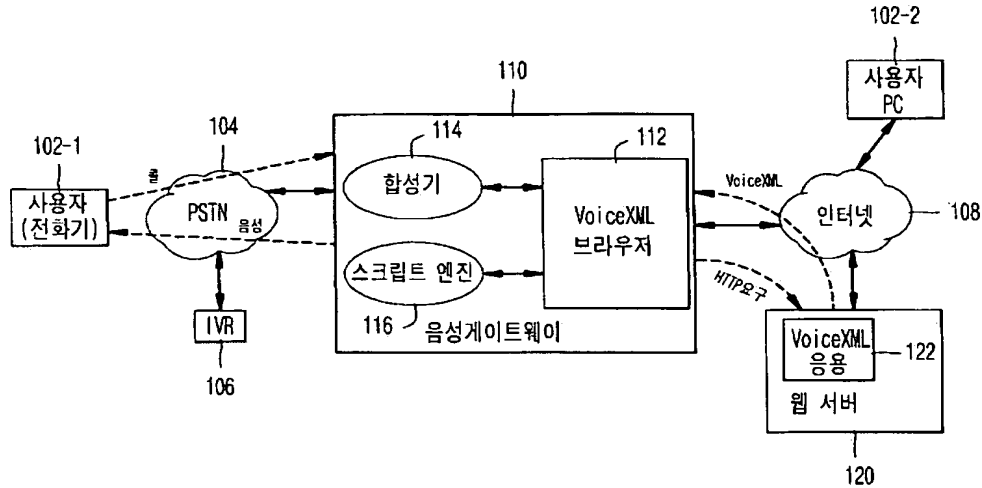
제6항 또는 제7항에 있어서,

상기 VoiceXML-to-XHTML+Voice 변환기는 상기 VoiceXML 트리를 상위 태그부터 하위 태그까지 모두 검색하면서, 태그를 확인하여 태그가 <menu>이면 XHTML의 <a> 태그로 변환하고, <grammar>이면 XHTML의 <input type=radio> 태그로 변환하며, 태그가 <form>이면 XHTML의 <form>을 XHTML 트리에 추가한 후 <form> 태그를 처리하는 것을 특징으로 하는 멀티모달 서비스 시스템.

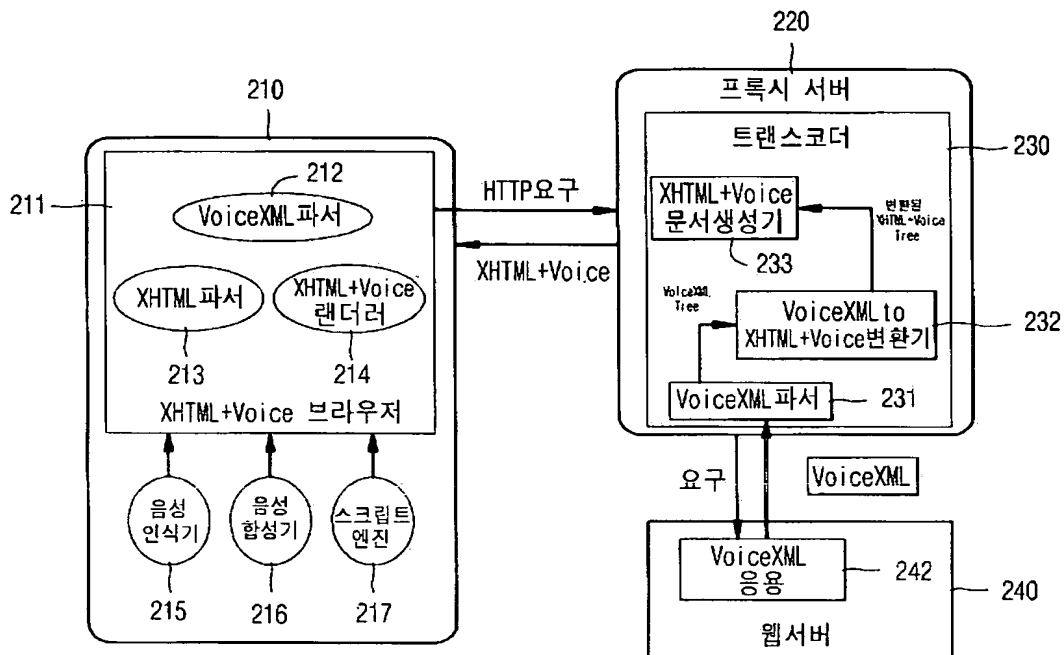


【도면】

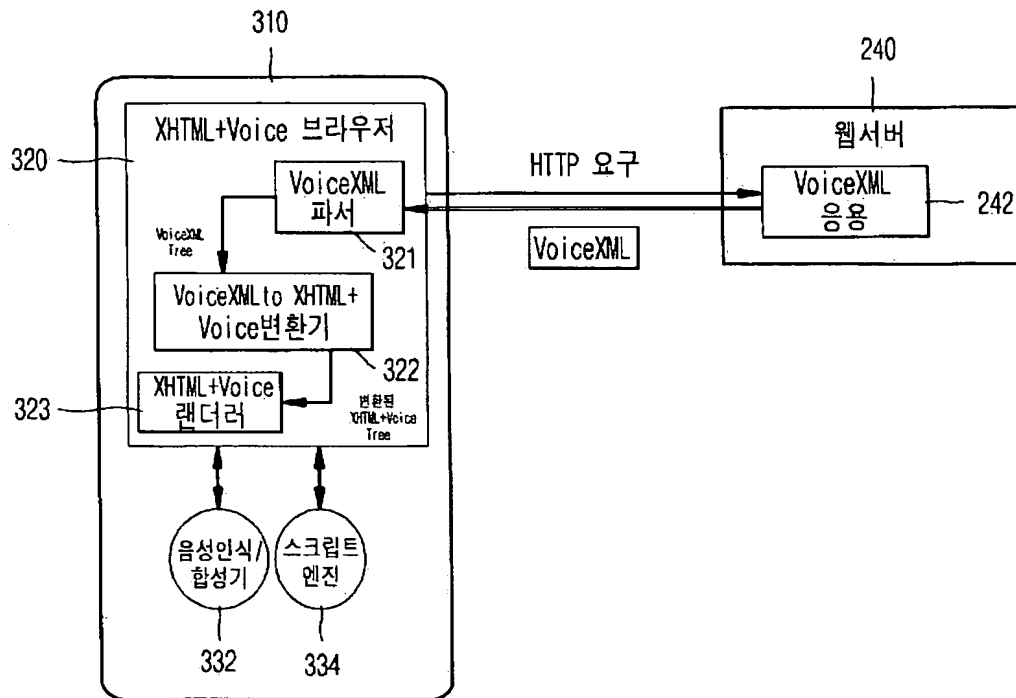
【도 1】



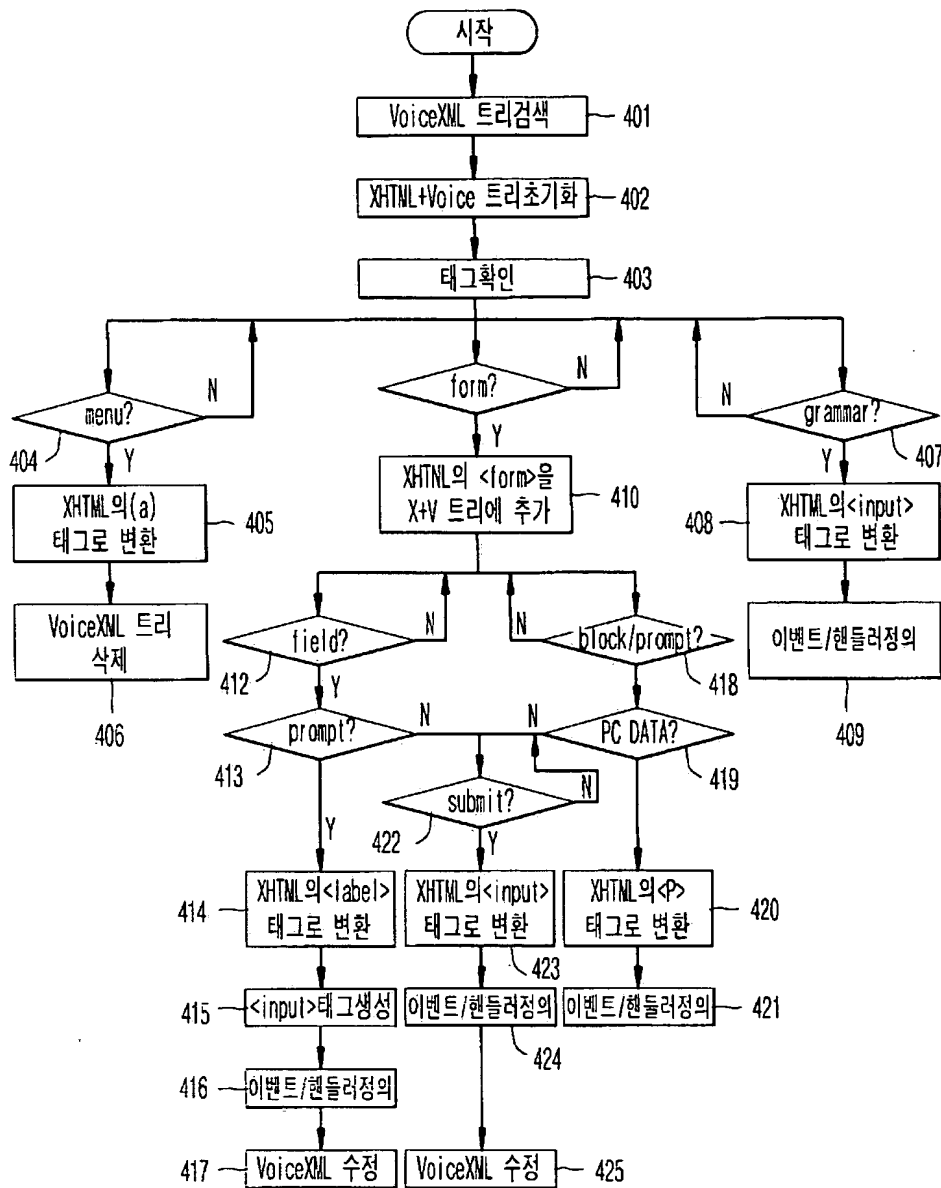
【도 2】



【도 3】

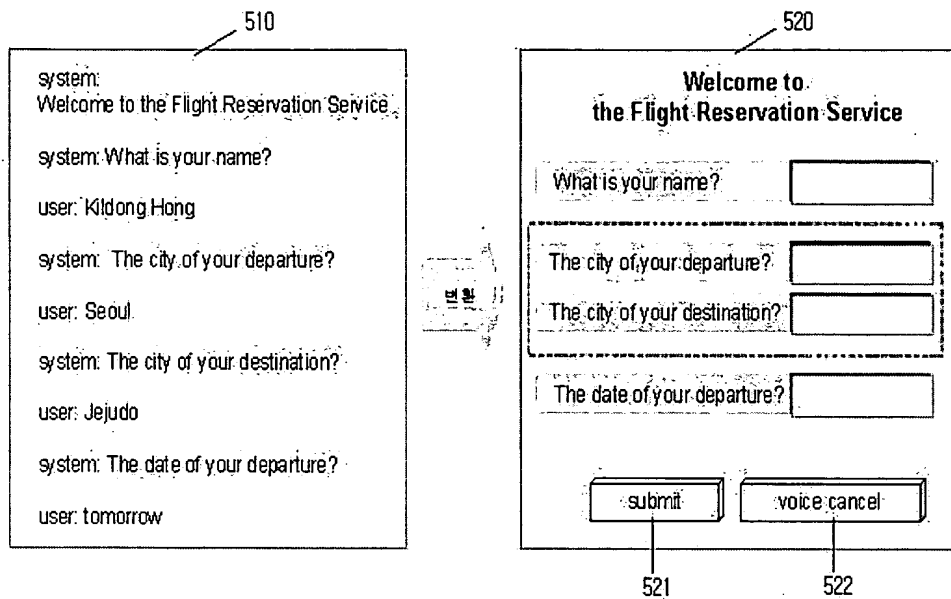


【도 4】

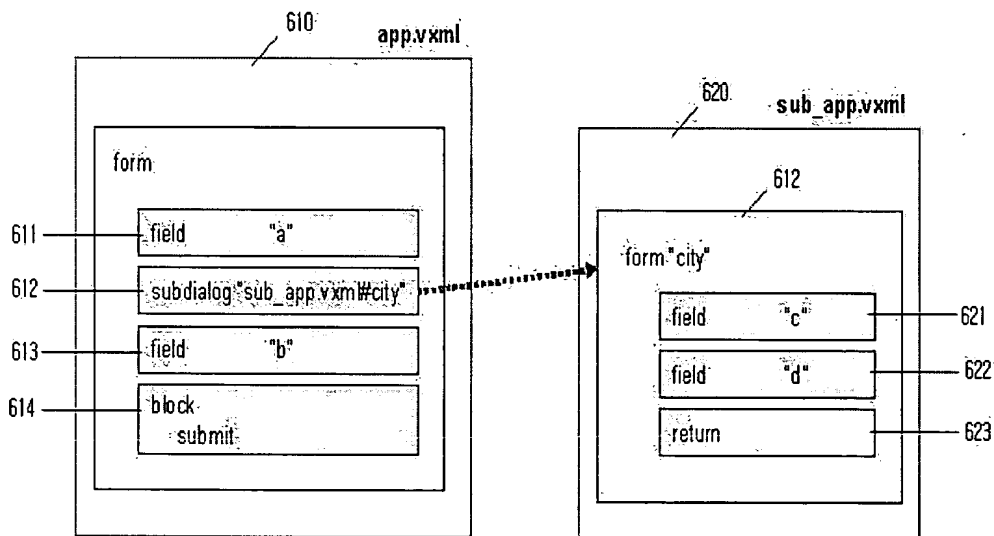




【도 5】

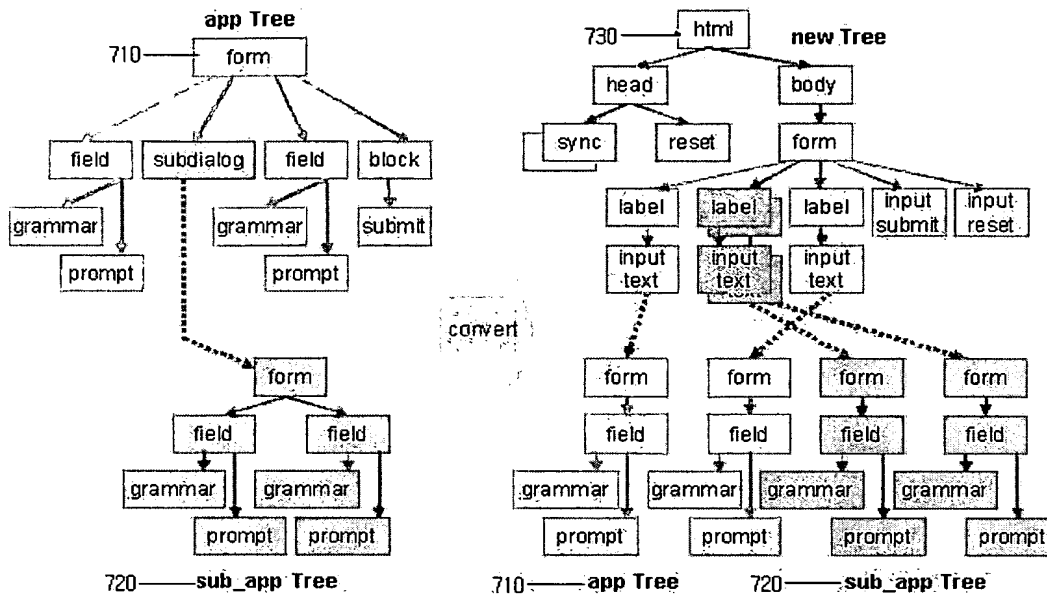


【도 6】





【도 7】



【도 8】

